



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :  C12N 9/90		A1	(11) International Publication Number: <b>WO 98/48006</b>  (43) International Publication Date: 29 October 1998 (29.10.98)		
(21) International Application Number: PCT/SE98/00703  (22) International Filing Date: 17 April 1998 (17.04.98)		(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).			
(30) Priority Data: 9701454-2 18 April 1997 (18.04.97) SE					
(71)(72) Applicants and Inventors: LINDAHL, Ulf [SE/SE]; Torgvägen 7, S-756 46 Uppsala (SE). LI, Jin-ping [SE/SE]; Reykjaviksgatan 51, S-752 63 Uppsala (SE).					
(74) Agent: AWAPATENT AB; P.O. Box 45086, S-104 30 Stockholm (SE).					
		Published <i>With international search report.</i>			
(54) Title: DNA SEQUENCE CODING FOR A MAMMALIAN GLUCURONYL C5-EPIMERASE AND A PROCESS FOR ITS PRODUCTION					
(57) Abstract					
An isolated or recombinant DNA sequence coding for a mammalian, including human, glucuronyl C5-epimerase or a functional derivative thereof capable of converting D-glucuronic acid (GlcA) to L-iduronic acid (IdoA); a recombinant expression vector comprising such DNA sequence; a host cell transformed with such recombinant expression vector; a process for the manufacture of a glucuronyl C5-epimerase or functional derivative thereof capable of converting GlcA to IdoA, comprising cultivation of a cell-line transformed with such recombinant expression vector, and a glucuronyl C5-epimerase or functional derivative thereof prepared by such process.					

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

DNA SEQUENCE CODING FOR A MAMMALIAN GLUCURONYL C5-EPIMERASE AND A PROCESS FOR ITS PRODUCTION

The present invention relates to an isolated or recombinant DNA sequence coding for a glucuronyl C5-epimerase capable of converting D-glucuronic acid to L-iduronic acid. The invention also relates to a process 5 for the manufacture of such epimerase.

Background of the invention

Heparin and heparan sulfate are complex, sulfated glycosaminoglycans composed of alternating glucosamine 10 and hexuronic acid residues. The two polysaccharides are structurally related but differ in composition, such that heparin is more heavily sulfated and shows a higher ratio of L-iduronic acid (IdoA)/D-glucuronic acid (GlcA) units (Kjellén, L. and Lindahl, U. (1991) Annual Review of Biochemistry 60, 443-475; Salmivirta, M., Lidholt, K. and Lindahl, U. (1996) The FASEB Journal 10, 1270-1279). Heparin is mainly produced by connective tissue-type mast cells, whereas heparan sulfate has a ubiquitous distribution and appears to be expressed by most cell types. The 20 biological roles of heparin and heparan sulfate are presumably largely due to interactions of the polysaccharides with proteins, such as enzymes, enzyme inhibitors, extracellular-matrix proteins, growth factors/cytokines and others (Salmivirta, M., Lidholt, K. and Lindahl, U. 25 (1996) The FASEB Journal 10, 1270-1279). The interactions tend to be more or less selective/specific with regard to carbohydrate structure, and thus depend on the amounts and distribution of the various sulfate groups and hexuronic acid units. Notably, IdoA units are believed to generally promote binding of heparin and heparan sulfate chains to proteins, due to the marked conformational flexibility of these residues (Casu, B., Petitou, M., Provasoli, M. and Sinay, P. (1988) Trends in Biochemical Sciences 13, 221-225).

Heparin and heparan sulfate are synthesized as proteoglycans. The process is initiated by glycosylation reactions that generate saccharide sequences composed of alternating GlcA and N-acetylglucosamine (GlcNAc) units 5 covalently bound to peptide core structures. The resulting  $(\text{GlcA}\beta 1,4-\text{GlcNAc}1,4-)_n$  disaccharide repeats are modified, probably along with chain elongation, by a series of enzymatic reactions that is initiated by N-deacetylation and N-sulfation of GlcNAc units, continues through 10 C-5 epimerization of GlcA to IdoA residues, and is concluded by the incorporation of O-sulfate groups at various positions. The N-deacetylation/N-sulfation step has a key role in determining the overall extent of modification of the polymer chain, since the GlcA C-5 epimerase 15 as well as the various O-sulfotransferases all depend on the presence of N-sulfate groups for substrate recognition. While the GlcNAc N-deacetylation and N-sulfation reactions are both catalyzed by the same protein, isolation and molecular cloning of N-deacetylase/N-sulfotrans- 20 ferase from different tissue sources implicated two distinct forms of the enzyme. The two enzyme types differ with regard to kinetic properties, and it has been suggested that they may be differentially involved in the biosynthesis of heparin and heparan sulfate.

25

Summary of the invention

The present invention provides for an isolated or recombinant DNA-sequence coding for a mammalian, including human, glucuronyl C-5 epimerase or a functional derivative thereof capable of converting D-glucuronic acid (GlcA) to L-iduronic acid (IdoA).

The invention also provides for a recombinant expression vector containing a transcription unit comprising a DNA sequence as described above, a transcriptional 30 promoter, and a polyadenylation sequence.

The invention also provides for a process for the manufacture of a glucuronyl C-5 epimerase or a functional

derivative thereof capable of converting D-glucuronic acid (GlcA) to L-iduronic acid (IdoA), comprising cultivation of a cell line transformed with the above recombinant expression vector in a nutrient medium allowing expression and secretion of said epimerase or functional derivative thereof.

Specific DNA sequences according to the invention are defined in appended claims 2, 3 and 4.

Furthermore, the invention provides for a host cell 10 transformed with such recombinant expression vector.

Finally, the invention covers a glucuronyl C-5 epimerase or a functional derivative thereof whenever prepared by the process outlined above.

15 Brief description of the appended figures and sequence listing

**Sequence listing: Nucleotide sequence and the predicted amino acid sequence of the C5-epimerase.** The predicted amino acid sequence is shown below the nucleotide sequence. The numbers on the right indicate the nucleotide residue and the amino acid residue in the respective sequence. The five sequenced peptides appear in **bold**. The N-terminal sequence of the purified protein is shown in **bold and italics**. The potential N-glycosylation sites (\*) are shown. The potential transmembrane region is underlined.

**Fig 1. In vitro transcription-translation.** The epimerase cDNA was inserted into a pcDNA3 expression vector and 30 linearized with XbaI at the 3'-end. It was then subjected to *in vitro* transcription-translation in a rabbit reticulocyte lysate system in the presence of [<sup>35</sup>S]methionine, as described in "Experimental Procedures". The translation product of epimerase cDNA (Epi) has a molecular 35 weight of ~50 kDa, by comparison with the LMW protein standard. A 118 kDa control sample of  $\beta$ -galactosidase

(C), expressed in the same system, is shown for comparison.

**Fig 2. Effect of the expressed C5-epimerase on N-deacetylated, N-sulfated capsular polysaccharide from E. coli K5.** Metabolically  $^3\text{H}$ -labeled K5 polysaccharide was N-deacetylated and N-sulfated, and was then incubated with (A) lysate of Sf9 cells infected with recombinant C5-epimerase; (B) lysate of Sf9 cells infected with recombinant  $\beta$ -glucuronidase. The incubation products were treated with  $\text{HNO}_2 / \text{NaBH}_4$ , and the resultant hexuronyl-anhydromannitol disaccharides were recovered and separated by paper chromatography. The arrowheads indicate the migration positions of glucuronosyl-anhydromannitol (GM) and iduronosyl-anhydromannitol (IM) disaccharide standards. For further information see "Experimental Procedures".

**Fig 3. Northern analysis of C5-epimerase mRNA expressed in bovine lung and mastocytoma cells.** Total RNA from each tissue/cell line was separated by agarose gel electrophoresis. A blot was prepared, probed with a  $^{32}\text{P}$ -labeled 2460-bp fragment of the epimerase cDNA clone, and finally exposed to X-ray film. (Kodak, Amersham). The arrow indicates the positions of molecular standards. For further information see "Experimental Procedures".

#### Detailed description of the invention

The present invention relates to DNA sequences coding for a mammalian glucuronyl C5-epimerase or a functional derivative thereof, such epimerase or derivative being capable of converting D-glucuronic acid (GlcA) to L-iduronic acid (IdoA). The term "mammalian" is intended to include also human varieties of the enzyme.

As used herein the definition "glucuronyl C5-epimerase or a functional derivative thereof" refers to enzymes which have the capability of converting D-

glucuronic acid to L-iduronic acid. Accordingly, the definition embraces all epimerases having such capability including functional variants, such as functional fragments, mutants resulting from mutageneses or other recombinant techniques. Furthermore, the definition is intended to include glycosylated or unglycosylated mammalian glucuronyl C5-epimerases, polymorphic or allelic variants and other isoforms of the enzyme. "Functional derivatives" of the enzyme can include functional fragments, functional fusion proteins or functional mutant proteins. Such epimerases included in the present invention can have a deletion of one or more amino acids, such deletion being an N-terminal, C-terminal or internal deletion. Also truncated forms are envisioned as long as they have the conversion capability indicated herein.

Operable fragments, mutants or truncated forms can suitably be identified by screening. This is made possible by deletion of for example N-terminal, C-terminal or internal regions of the protein in a step-wise fashion, and the resulting derivative can be analyzed with regard to its capability of the desired conversion of D-glucuronic acid to L-iduronic acid. If the derivative in question operates in this capacity it is considered to constitute a functional derivative of the epimerase proper.

Examples of useful epimerases are proteins having the sequence as shown in the sequence listing or substantially as shown in the sequence listing and functional portions thereof.

30

#### EXPERIMENTAL PROCEDURES

*Peptide Purification and Sequencing -* The 52 kDa epimerase protein (~1 $\mu$ g), purified from a detergent extract of bovine liver by chromatography on O-desulfated heparin-Sepharose, Red-Sepharose, Phenyl-Sepharose, and Concanavalin A-Sepharose (Campbell, P., Hannesseon, H.H., Sandbäck, D., Rodén, L., Lindahl, U. and Li, J.-p. (1994)

J Biol Chem 269, 26953-26958), was subjected to direct N-terminal sequencing using a model 470A protein sequenator (Applied Biosystems) equipped with an on-line 120 phenylthiohydantoin analyzer (Tempst, P., and Riviere, L. 5 (1989) Anal. Biochem. 183, 290-300). Another sample (~1 $\mu$ g) was applied to preparative (12%) SDS-PAGE and was then transferred to a PVDF membrane. After staining the membrane with Coomassie Blue, the enzyme band was excised. Half of the material was submitted to direct N-terminal sequence analysis, whereas the remainder was digested with Lys-C (0.0075 U; Waco) in the presence of 1% RTX-100/10% acetonitrile/100mM Tris-HCl, pH 8.0. The generated peptides were separated on a reverse phase C4-10 column, eluted at a flow rate of 100  $\mu$ l/min with a 6-ml 15 10-70% acetonitrile gradient in 0.1% trifluoroacetic acid, and detected with a 990 Waters diode-array detector. Selected peptides were then subjected to sequence analysis as described above.

Probes for Screening - Total RNA was extracted from 20 bovine liver according to the procedures of Sambrook et al. (1989). Single-stranded cDNA was synthesized by incubating ~5  $\mu$ g of bovine liver total RNA (denatured at 65°C, 3 min) with a reaction mixture containing 1 unit RNase inhibitor (Perkin-Elmer Corp.), 1 mM of each dNTP, 25 5  $\mu$ M random nucleotide hexamer and 1.25 units of murine leukemia virus reverse transcriptase (Perkin-Elmer Corp.) in a buffer of 10 mM Tris-HCl, pH 8.3. The mixture was kept at 42°C for 45 min and then at 95°C for 5 min. Degenerated oligonucleotide primers were designed based on 30 the amino-acid sequence determined for one of the internal peptides derived from the purified epimerase (Table I). Single-stranded bovine liver cDNA was applied to PCR together with 100 pmols of primers 1 (sense) and 3 (antisense), in a total volume of 100  $\mu$ l containing 1 $\mu$ l 35 of 10% Tween 20, 6 mM MgCl<sub>2</sub>, 1 mM of each dNTP, and 2.5 units Taq polymerase (Pharmacia Biotech) in a buffer of 10 mM Tris-HCl, pH 9.0. The reaction products were sepa-

rated on a 12% polyacrylamide gel. A ~100-bp band was cut out from the gel and reamplified using the same PCR conditions. After an additional polyacrylamide gel electrophoresis, the product was isolated and sequenced ,

5 yielding a 108-bp sequence. This PCR product was subcloned into a pUC119 plasmid. The DNA fragment cleaved from the plasmid was labeled with [<sup>32</sup>P]dCTP (DuPont NEN) using a Random Primed DNA Labeling Kit (Boehringer Mannheim).

10        *Screening of cDNA Library* - A bovine lung cDNA library constructed in a lgt10 vector (Clontech) was screened with the 108-bp PCR fragment as hybridizing probe. The nitrocellulose replicas of the library plaques were prehybridized in 6xSSC, 5xDenhardt's solution containing 0.1% SDS and 0.1 mg/ml denatured salmon DNA for 2 hours at 65°C. Hybridization was carried out at 42°C in the same solution containing <sup>32</sup>P-labeled probe for 16-18 hours. The filters were washed two times with 2x SSC, 0.5% SDS and two times with 0.5x SSC, 1% SDS at the same

15 temperature. The library was repeatedly screened twice under the same conditions. Finally, the entire cDNA phage library was subjected to PCR amplification using lgt10 forward and reverse primers (Clontech) with a epimerase cDNA specific primer (5'-GCTGATTCTTTCTGTC-3').

20        *Subcloning and Sequencing of cDNA Inserts* - cDNA inserts, isolated by preparative agarose gel electrophoresis (Sambrook et al., 1989) after EcoRI restriction cleavage of recombinant bacteriophage DNA, were subcloned into a pUC119 plasmid. The complete nucleotide sequence was determined independently on both strands using the dideoxy chain termination reaction either with [<sup>35</sup>S]dATP and the modified T7 DNA polymerase (Sequenase version 2.0 DNA Sequencing Kit; U. S. Biochemical Corp.) or the ALF™ System (Pharmacia Biotech). DNA sequences were compiled and analyzed using the DNASTARTM program (Lasergene).

25        *Polyclonal Antibodies and Immunodetection*- A peptide corresponding to residues 77 - 97 of the deduced epime-

rase amino-acid sequence was chemically synthesized (Åke Engström, Department of Medical and Physiological Chemistry, Uppsala University, Sweden), and was then conjugated to ovalbumin using glutaraldehyde (Harlow, E. and Lane, 5 D. (1989) in Antibodies: A Laboratory Manual, pp 78-79, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). A rabbit was immunized with the peptide conjugates together with Freund's adjuvant. After 6 boosts (each with 240 µg conjugated peptide) blood was collected and the serum recovered. The antibody fraction was further purified on a 10 Protein A-Sepharose column (Pharmacia Biotech), and used for immunoblotting.

Samples of GlcA C5-epimerase were separated under denaturing conditions by 12% SDS-PAGE, and were then 15 transferred to a nitrocellulose membrane (Hybond™ ECL). ECL immunoblotting was performed according to the protocol of the manufacturer (Amersham). Briefly, the membrane was first treated with blocking agent, then incubated with purified antibody, and finally incubated with the 20 peroxidase labeled anti-rabbit antibody. After adding the ECL reagent, the light emitted by the chemical reaction was detected by exposure to Hyperfilm™ ECL for 30-60 sec.

*Northern Blot Hybridization* -- Bovine liver and lung total RNA was prepared according to Sambrook, J., 25 Fritsch, E.F. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual, Cold Spring Harabor Laboratory, Cold Spring Harbor NY), and mouse matocytoma (MCT) total RNA was extracted from a tumor cell line (Montgomery, R.I., Lidholt, K., Flay, N.W., Liang, J., Verter, B., 30 Lindahl, U. and Esko, J.D. (1992) PNAS **89**, 11327-113331) as described by Chomczynski and Sacchi (1987). Total RNA from each tissue (~20 µg samples) was denatured in 50% formamide (v/v), 5% formaldehyde, 20 mM Mops buffer, pH 7.0, at 65 °C for 5 min. The denatured RNA was separated 35 by electrophoresis in 1.2% agarose gel containing 5% formaldehyde (v/v), and was then transferred to a Hybond N<sup>+</sup> nylon membrane (Amersham). The RNA blot was pre-hybri-

dized in ExpressHyb Hybridization Solution (Clontech) at 65 °C for 1 h, and subsequently hybridized in the same solution with-a [<sup>32</sup>P]dCTP-labeled DNA probe (a 2460 bp fragment including the 5'-end of the cDNA clone; see the 5 sequence listing). The membrane was washed in 2x SSC, 0.5% SDS at the same temperature for 2 x 15 min and in 0.5x SSC, 0.5 % SDS for 2 x 15 min. The membrane was exposed to a Kodak X-ray film at -70°C for 24h.

*In Vitro Translation* - The 3-kb GlcA C5-epimerase 10 clone, inserted in a pcDNA3 expression vector (Invitrogen) was linearized at the 3'-end by restriction enzyme *Xba*I. *In vitro* translation was carried out with a Linked T7 transcription-translation system (Amersham) according to the instructions of the manufacturer. The corresponding mRNA generated by incubation of 0.5 µg linearized plasmid DNA with a T7 polymerase transcription mix (total volume, 10 µl; 30°C; 15 min) was mixed with an optimized rabbit reticulocyte lysate containing 50µCi [<sup>35</sup>S]methionine (total volume, 50 µl), and further incubated at 30 °C for 1 h. A sample (5 µl) of the product 15 was subjected to 12% SDS-PAGE. The gel was directly exposed to a Kodak X-ray film. After exposure, the applied protein molecular standards ( LMW Molecular Calibration Kit, Pharmacia Biotech) were visualized by staining the 20 gel with Coomassie Blue.

*Expression of the GlcA C5-Epimerase* - The GlcA C5-epimerase was expressed using a BacPAK8™ Baculovirus Expression System (Clontech), according to the instructions by the manufacturer. Two oligonucleotides, one at the 5'- 30 end of the cDNA clone (1-17 bp, sense) and the other at the 3'-end of the coding sequence (1387-1404 bp, antisense), were used to PCR amplify the coding sequence of the C5-epimerase cDNA clone. The resulting fragment was cloned into the BacPAK8 vector. Sf9 insect cells, maintained in Grece's Insect Medium (GibcoBRL) supplemented 35 with 10% fetal calf serum and penicillin/streptomycin, were then cotransfected by the C5-epimerase construct

along with viral DNA. Control transfections were performed with constructs of a  $\beta$ -glucuronidase cDNA construct included in the expression kit, and a mouse cDNA coding for the GlcNAc N-deacetylase/N-sulfotransferase implicated in heparin biosynthesis (Eriksson, I., Sandbäck, D., Ek, B., Lindahl, U. and Kjellén, K. (1994) J.Biol. Chem. **269**, 10438-10443; Cheung, W.F., Eriksson, I., Kusche-Gullberg, M., Lindahl, U. and Kjellén, L. (1996) Biochemistry **35**, 5250-5256). Single plaques of each co-transfected recombinant were picked and propagated. Two Petri dishes (60-mm) of Sf9 cells were infected by each recombinant virus stock and incubated at 27°C for 5 days. The cells from one dish were used for total RNA extraction and Northern analysis performed as described above. Cells from the other dish were lysed in a buffer of 100 mM KCl, 15 mM EDTA, 1% Triton X-100, 50 mM HEPES, pH 7.4, containing 1mM PMSF and 10 $\mu$ g/ml pepstatin A. Supernatants of cell lysates as well as conditioned media were analyzed for epimerase activity. Protein contents of the cell lysates were estimated by the method of Bradford (1976) or by the BCA reagent procedure (Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985) Anal. biochem **150**, 76-85).

*Demonstration of GlcA C5-epimerase activity - Epimerase activity was assayed using a biphasic liquid scintillation counting procedure, essentially as described by Campbell et al. (1994) above. The reaction mixtures, total volume 55  $\mu$ l, contained 25  $\mu$ l cell lysate or medium, 25  $\mu$ l of 2x epimerase assay buffer (20 mM HEPES, 30 mM EDTA, 0.02% Triton X-100, 200 mM KCl, pH 7.4) and 5  $\mu$ l of substrate (10,000 cpm  $^3$ H). The substrate was a chemically N-deacetylated and N-sulfated polysaccharide, obtained from *E. coli* K5 according to the procedure of Campbell et al. (1994), except that D-[5- $^3$ H]glucose was substituted for D-[1- $^3$ H]glucose.*

Enzymatic conversion of D-glucuronic to L-iduronic acid was demonstrated using the metabolically  $^{1-3}\text{H}$ -labeled substrate (N-deacetylated, N-sulfated capsular polysaccharide from *E. coli* K5) and the analytical procedure described by Campbell et al. (1994). A sample (~20 µg; 200,000 cpm of  $^3\text{H}$ ) of the modified polymer was incubated with 250 µl of cell lysate in a total volume of 300 µl epimerase assay buffer at 37°C for 6 hours. The incubation was terminated by heating at 100°C for 5 min.

The sample was mixed with 50µg of carrier heparin and reacted with nitrous acid at pH 1.5 (Shively, J., and Conrad, H.E. (1976) *Biochemistry* **15**, 3932-3942), followed by reduction of the products with NaBH<sub>4</sub>. The resultant hexuronyl-anhydromannitol disaccharides were recovered by gel chromatography on a column (1 x 200 cm) of Sephadex G-15 in 0.2 M NH<sub>4</sub> HCO<sub>3</sub>, lyophilized, and subjected to paper chromatography on Whatman No 3MM paper in ethyl acetate/acetic acid/water (3:1:1).

RESULTS

*Generation of a Probe and Screening of cDNA Library*

- Amino acid sequence data for the ~52 kDa protein were obtained by digesting highly purified epimerase with lysine-specific protease, followed by separation of the generated peptides on a reverse phase column. The five most prominent peptides were isolated and subjected to amino-acid sequencing (Table I). One of the peptides (peptide 1) was found to correspond to the N-terminal sequence of the native protein. The sequence of the largest peptide obtained (peptide 5 in Table I), was used to design two sense and one antisense degenerate oligonucleotide primers, as shown in Table I. A DNA probe was produced by PCR using primers 1 and 3 with bovine liver cDNA as template. The resultant ~100 bp DNA fragment was purified by polyacrylamide gel electrophoresis, reamplified using the same primers, and finally isolated by electrophoresis. The identity of the product was ascertained by

"nested" PCR, using primers 2 and 3, which yielded the expected ~60 bp fragment (data not shown). Moreover, sequencing of the larger (108 bp) DNA fragment gave a deduced amino-acid sequence identical to that of the isolated peptide (Table I).

The 108-bp PCR product was labeled with [<sup>32</sup>P]dCTP and used for screening of a bovine lung lgt10 library. One hybridizing clone, containing a 3-kb insert, was identified. Repeated screening of the same library yielded two additional positive clones, both of which were of smaller size. Subsequent sequencing showed both of the latter clones to be contained within the 3.0-kb species (data not shown). The 3-kb clone was sequenced through both strands and was found to contain altogether 3073 bp; an additional 12-bp sequence was added at the 5'-end through characterization of a separate clone obtained by PCR amplification of the phage library (see "Experimental Procedures").

*Characterization of cDNA and Predicted Protein*

Structure - The total cDNA sequence identified, in all 3085 bp, contains an open reading frame corresponding to 444 amino-acid residues (the sequence listing). Notably, the coding region (1332 bp) is heavily shifted toward the 5'-end of the available cDNA, and is flanked toward the 3'-end by a larger (1681 bp) noncoding segment. The deduced amino-acid sequence corresponds to a 49,905 dalton polypeptide. All of the five peptides isolated after endo-peptidase digestion (Table I) were recognized in the primary structure deduced from the cDNA (the sequence listing). One of these peptides (peptide 1) is identical to the N-terminus of the isolated liver protein. This peptide was found to match residues 74 - 86 of the deduced polypeptide sequence. The enzyme isolated from bovine liver thus represents a truncated form of the native protein.

Generation of mRNA from an expression vector inserted with the 3-kb cDNA clone, followed by incubation

of the product with rabbit reticulocyte lysate in the presence of [<sup>35</sup>S]methionine, resulted in the formation of a distinct labeled protein with an estimated M<sub>r</sub> of ~50kDa (Fig. 1). This product was recognized in immunoblotting  
5 (data not shown) by polyclonal antibodies raised against a synthetic peptide corresponding to residues 77 - 97 (see the sequence listing) of the deduced amino-acid sequence. The same antibodies also reacted with the isolated ~52 kDa bovine liver protein (data not shown).  
10 These observations establish that the 3-kb cDNA is derived from the transcript that encodes the isolated ~52 kDa bovine liver protein.

The cDNA structure indicates the occurrence of 3 potential N-glycosylation sites (the sequence listing).  
15 Sugar substituents may be important for the proper folding and catalytic activity of the enzyme, since the protein expressed in bacteria (which also gave a strong Western signal towards the polyclonal antibodies raised against the synthetic peptide; data not shown) was devoid  
20 of enzymatic activity. A potential transmembrane region is underlined in the sequence listing. The predicted protein contains two cystein residues, only one of which occurs in the isolated (truncated) protein. Since NEM was inhibitory to epimerase activity (data not shown), this  
25 single cystein unit may be essential to the catalytic mechanism.

*Functional Expression of the GlcA C5-Epimerase* - A variety of expression systems were tested in attempts at generating the cloned protein in catalytically active  
30 form. A protein obtained by in vitro translation using a rabbit reticulocyte lysate system (see Fig. 1) showed no detectable epimerase activity. A construct made by inserting the 3-kb cDNA into a pcDNA3 vector (Invitrogen) failed to induce mRNA formation (or translation) in any  
35 of the cell lines tested (human embryonic kidney (293), COS-1 or CHO cells) (data not shown). We also attempted to express the enzyme in a bacterial pET system

(Novagen). The transformed bacteria yielded appreciable amounts of immunoreactive protein which, however, lacked detectable enzyme activity (data not shown).

Cotransfection of epimerase recombinant with baculovirus into Sf9 insect cells resulted in the generation of abundant GlcA C5-epimerase activity (Table II). In two separate experiments, the lysates from cells infected with the same epimerase recombinant virus stock showed >10-fold higher enzyme activities, on a mg protein basis, than the corresponding fractions from cells infected with control recombinant virus stock. The conditioned media of cells infected with epimerase recombinant showed 20- 30-fold higher enzyme activities than the corresponding fractions from cells infected with control plasmid virus stock. Transfections with cDNA encoding other enzymes, such as a  $\beta$ -glucuronidase, or the mouse mastocytoma GlcNAc N-deacetylase/N-sulfotransferase involved in heparin biosynthesis (Eriksson et al., 1994), did not significantly increase the epimerase activity beyond control levels. Notably, the higher  $^3\text{H}_2\text{O}$  release recorded for control samples as compared to heat-inactivated expressed enzyme (Table II) suggests that the insect cells constitutively produce endogenous C5-epimerase.

The polysaccharide substrate used for routine assays of epimerase activity was obtained by chemically N-deacetylation and N-sulfation of the capsular polysaccharide [(GlcA $\beta$ 1,4-GlcNAc1,4)n] of E. coli K5 that had been grown in the presence of [5- $^3\text{H}$ ]glucose. The data in Table II thus reflect the release of  $^3\text{H}_2\text{O}$  from 5- $^3\text{H}$ -labeled GlcA units in the modified polysaccharide, due to enzyme action (Jacobsson, I., Bäckström, G., Höök, M., Lindahl, U., Feingold, D.S., Malmström, M., and Rodén, L. (1979) *J.Biol. Chem.* 254, 2975-2982; Jacobsson, I., Lindahl, U., Jensen, J.W., Rodén, L., Prihar, H. and Feingold, D.S. (1984) *Journal of Biological Chemistry* 259, 1056-1064). More direct evidence for the actual conversion of GlcA to IdUA residues was obtained by incubating the expressed

enzyme with an analogous substrate, obtained following incubation of the bacteria with [ $1-^3\text{H}$ ]glucose. This substrate will retain the label through the epimerization reaction, and can therefore be used to demonstrate the formation of IdoA-containing disaccharide units. Following incubation with the recombinant epimerase, 21% of the hexuronic acid residues was converted to IdoA, as demonstrated by paper chromatography of disaccharide deamination products (Fig. 2). The composition of the incubated polysaccharide thus approached the equilibrium ratio of IdoA/GlcA, previously determined to ~3/7<sup>1</sup>).

*Northern Analysis* -Total RNA, from bovine liver, lung, and mouse mastocytoma, were analysed by hybridization with a 2460-bp DNA fragment from epimerase cDNA clone as a probe. Both bovine liver and lung gave identical transcription patterns, with a dominant transcript of ~9 kb and a weak ~5 kb band (Fig. 3). By contrast, the mastocytoma RNA showed only the ~5 kb transcript.

It is to be noted that the present invention is not restricted to the specific embodiments of the invention as described herein. The skilled artisan will easily recognize equivalent embodiments and such equivalents are intended to be encompassed in the scope of the appended claims.

Table I  
Peptide and primer sequences

<b>A. N-terminal sequences of isolated C5-epimerase</b>			
1.	PNDWXVPKGCFMA	(free solution)	
2.	PXDWTVPKGXF	(band excised from PVDF-membrane)	
<b>B. Peptide sequences</b>			
1.	PNDXTVPK		
2.	XXIAPETSEGXSQL		
3.	GGWPIMVTRK		
4.	FLSEQHGV		
5.	<u>KAMLPLYDTGSGTIYDLRHFMGLGIAPNLAXWDYHTT</u>		
	primer 1 (sense)	primer 2 (sense)	primer 3 (antisense)
<b>C. Primer<sup>a</sup></b>			
		Degeneracy	
1 (S)	5'-cc gaattcAARGCNATGYTNCCNYT-3' <sup>b</sup>	384	
2 (S)	5'-cc gaattcGAYYTNCAGTCATG-3'	288	
3 (AS)	5'-cc ggatccGTNGTRTGRTARTCCCA-3'	32	

<sup>a</sup> (R, A or G; Y, T or C; M, C or A; N, A or C or G or T)

<sup>b</sup> (cc, clamp; gaatcc, EcoRI restriction site; ggatcc, BamHI restriction site)

Table II  
Expression of HexA C5-epimerase in Sf9 cells

Sf9 cells ( $1 \times 10^6$  in 4 ml medium) were seeded in 60-mm Petri dishes and incubated for three hours at 27°C. After the cells were attached, the medium was removed, and 200 µl of recombinant virus stock was added to infect the cells at room temperature for 1h. The virus suspension was aspirated and 4 ml of medium was added to each dish. The cells were incubated at 27 °C for 5 days. The medium was transferred into a steril tube and centrifuged. The cells were collected, washed twice with PBS and lysed with 300 µl of homogenization buffer as described under "Experimental Procedures". Aliquots (25 µl) of cell lysate and medium were assayed for epimerase activity. The activity is expressed as release of  $^3\text{H}$  from K5 polysaccharide per hour. The data is mean value of three independent assays.

Construct	Epimerase Activity	
	Cell lysate (cpm/mg/h)	Medium (cpm/ml/h)
HexA C5-Epimerase-1	$102670 \pm 5540$	$45200 \pm 1770$
HexA C5-Epimerase-2	$123270 \pm 4660$	$52610 \pm 810$
HexA C5-Epimerase-1 (heat-inactivted)	240	610
N-Deacetylase/sulfotransferase	$9520 \pm 620$	$1350 \pm 280$
$\beta$ -Glucuronidase	$8460 \pm 1270$	$1610 \pm 440$
BacPAK plasmid	$5150 \pm 880$	$2820 \pm 690$
Neo	$7250 \pm 370$	$550 \pm 120$

## SEQUENCE LISTING

TCCAACCTGAAATTCTCATACCTATTCCTAAAGCTATGCCACAGAGAGCCCCCTATCAACCT GATGGTGTGTTATGICCTTGAAGGCTACAATIGCGAAGTCCGAGACAGAGTCAGTCAGTC M S F E G Y N V E V R D R V K C	60 120 <b>16</b>
ATAAGTCCCCTTGAACGGTGTACCTTTATCTACACAGTGGGGACCTCAAGGCTATTTCTAC I S G V E G V P L S T O W G P O G Y F Y	180 <b>36</b>
CCAATOCAGATTGCCACAGTATGGGTTAACGCAAGAACAGGGACAAAAACACCAAGCCAAATGAC P I O I A O Y G L S H Y S K N L T E K P	240 <b>56</b>
CCTCATATAGAGGTATATGAAACACCGAGAACAGGGACAAAAACACCAAGCCAAATGAC P H I E V Y E T A E D R D K N S K P N D	300 <b>76</b>
TOGACTGTGCCAACGGCTGCCTTATGGCTAGTGTGGCTGATAAGTCAGATTACCAAT W T V P K G C F M A S V A D K S R F T N	360 <b>96</b>
GTTAACAGTCATTCTCCAGAAACCAGTGAAGGTGTATCCCTCCAATGGGAACACA V K Q F I A P E T S E G V S L Q L G N T	420 <b>116</b>
AAAGATTITATTATTTCATTTCATGACCTCAAGTCTTAACAAATGGAAGGGTGTCIGGTT K D F I I S F D L K F L T N G S V S V V	480 <b>136</b>
CCTGGAGACGACAGAAAAGAACATCAGCTCTCACTGTACATTATGTCCTCAAATACCCAGCTA L E T T E K N Q L F T V H Y V S N T Q L	540 <b>156</b>
ATIGCTTTAAAGAAAGAGACATATACTATGGCATGGGCCCAGAACATCATGGACACA I A F K E R D I Y Y G I G P R T S W S T	600 <b>176</b>
GTTACCCGGGACCTGGTCACTGACCTCACGGAAAGGAGTGGGCTTCCAAACACAAAAGCT V T R D L V T D L R K G V G L S N T K A	660 <b>196</b>
GTCAGCCAACAAGAATAATGCCAAAGAACGGTGGTTAGGTGATTCGAAAGCGAACGGC V K P T R I M P K K V V R L I A K G K G	720 <b>216</b>
TICCTTGACAACATTACCATCTCTACCAAGGCCACAGGGTGGCTGGCTTCTGGCTGGCAGT F L D N I T I S T T A H M A A F F A A S	780 <b>236</b>
GACTGGCTGGTGGAGAACCAAGATGAGAAAGGGGGCTGGCGATTATGGTGACCCGTAAG D W L V R N Q D E K G G W P I M V T R K	840 <b>256</b>
TTAGGGGAACCCCTCAAGCTTTAGAGCCACGGGGTACTCGGCCATGGCCAAGGCCAA L G E G F K S L E P G W Y S A M A Q G Q	900 <b>276</b>
GOCATTCTCACATTAGTCAGGCCCTATCTCTTAACAAAGACCATATACTCTCAATTCA A I S T L V R A Y L L T K D H I F L N S	960 <b>296</b>
GCCTTAAGGAAACAGGCCCTTACAAGTCTCTGGCTGGTCAAGGCTGIG A L R A T A P Y K F L S E Q H G V K A V	1020 <b>316</b>
TTATGAAATAACATGACTGGTATGAAGAATATCCAACCTACACCTAGCTCTTTGTTTA F M N K H D W Y E E Y P T T P S S F V L	1080 <b>336</b>
AAAGGCTTTATGATCTCTTAATTGGCTGTATGACTAAAGAAACTGCAGGGAAAAA N G F M Y S L I G L Y D L K E T A G E K	1140 <b>356</b>

CTCCCGAAAGAAGCGAGGCTCTGTATGACGGTCCCATTGGPATCCTTAAGGCCATCTC	1200
L G K E A R S L Y E R G M E S L K A M L	<b>376</b>
CCCTTGTAAGACACTCGCTCAAGGAPACCATCTATGACCTCCGGCACTTCATGCTTGGCATT	1260
P L Y D T G S G T I Y D L R H F M L G I	<b>396</b>
GCCCCCAACCTGGCCCCCTCGGACTATCACACCACCATCAATCAACTGGAGCTCCIT	1320
A P N L A R W D Y H T T H I N Q L Q L L	<b>416</b>
ACGACCATTTGATGAGGCCCCAATCTTCAAAGAATTGTCAAGAGGTGGAAGAGGCTACCTT	1380
S T I D E S P I F K E F V K R W K S Y L	<b>436</b>
AAAGCCAGCGGGCAAACACAACTAGACCTCAGAACAAAATCTACGTCAAGCTCTCC	1440
K G S R A K H N	<b>444</b>
TGTACACAGAAACTAGACCTCTGTGTCAAGCAGACCATAGGCACATTTAAAAGGCTGTA	1500
TAATCAAGTTTGTGGATTACATCAAAGTGATAATGATCCTTAAACCGAGTCCTCTGAG	1560
ATAATTGCAATTCAATGGTTAGTGTAGAATGTOGATGGCATTTATACCGAGAAAGTG	1620
TTTAGTCAGTGGGCTGAATGAAGAATGTTAACCTGGCTCGCTATCACCCCTGTCAGTT	1680
CCACACGGTAGTCAGTTCCTCGATTGGAAAGACAATGGTAAGTACCTCTGATGCC	1740
ACCTGTCACGCACTTGCTGAAACTTAGTATGGCCCTCTTTAAAATGGTTTATTTAT	1800
GTTTAATGTTGAAAGCAGACTTAAAAAAATATGCTTAAACAGTAAATAATGTAATT	1860
GTAGCCCTGATAGTGTGACTGTGCAACTTTAAAATGATTTCCTTCTTCTATAAAATT	1920
TCCTACGGGTGGATGACCAATTGTTGTTGTTGTCAGTTGTTATATATGGAGAATATT	1980
TGAATTATAGGTTGCTGAACTGTTAAACACAAACAGTTCAGGCTTCAC	2040
AGTTATATAATGTAAGCACAACCTAAATGAAACTTGTGACTGACAAGAAAATTACAAA	2100
CAGAACAAAATGTTATGTTTATGAAACTATCTACAATCAGTAAAGATTGATAATC	2160
AGTATAACCCCTCTGTACCCCCATTGGGGGGCTCTTGGCACTATCTCAAATTG	2220
TATTCATTTCAAGACTACACTTGAGAGTTTGCTATTTCGGGGGACATTGGGGACA	2280
TTGGGAAATTACTATAAACCTAGATTGATGAGGGTAGTAAGTTAAAGCCA	2340
CTACCACTGCCCTTCTCTAGATTCTTTCCTTAAAGAAAATATTAGTCAGATAITA	2400
TAAGGATTGTGAGATTCTTCTACTTACGATCATCTTGGCTACACCTTCCAAAC	2460
TATTGATGTCACAAAATACATAGTTTGTTGTTGTAACCTTCAAACTTTCTGGGTTT	2520
TCTTGGCAGTTTAAATTAAATTATTCACCTCTGGATAAAAGTGATGCTACTATAT	2580
TAGCTGTCACATGTGTAATCAGACCTTATTGGTTTATATCCACATACCTCACATAA	2640
ATAGGCATCATAGCCCTCACACCCCTGGGAGGTCCTGCTAGGACTTACCCAGTAGGT	2700
AGAATCTGAGGAGGTGATTTCCTGCTCTGTTAGTGTATGACAATACAGTAAATCA	2760
ATACAATAACTTATACAGATTGAAATACGAGATCGGTACTTTCAGAGGACTGAGTCG	2820
ACACACGCTGCTGAGTGTGTTGCTGACCTGATGAAATCCACATCAAGAGGGAGGTG	2880
OCTGCGCTGCCACTOCATCTGGCTGGACTTATGCTACCAACACCACTGAGAAATGCCAA	2940
ATCCATACTGCCCTTGTGACACATATAATGTTGATCAGAACCGGAACCAAGACCTACA	3000
ACTGACATGAACTGTTAGTCACTAAGTGTGTCACACTCTTGTGACCTCATAGACTGT	3060
AGCGCGCGAGGCTCTTGTCCATG	3085

CLAIMS

1. An isolated or recombinant DNA sequence coding for a mammalian, including human, glucuronyl C5-epimerase or a functional derivative thereof capable of converting 5 D-glucuronic acid (GlcA) to L-iduronic acid (IdoA).
2. A DNA sequence according to claim 1 constituted by a nucleotide sequence comprising nucleotide residues 1 to 1404, inclusive, as depicted in the sequence listing.
3. A DNA sequence according to claim 2 constituted 10 by a nucleotide residue comprising nucleotide residues 73 to 1404, inclusive, as depicted in the sequence listing.
4. A DNA sequence according to claim 2 constituted by a nucleotide residue comprising nucleotide residues 1 to 1404, inclusive, as depicted in the sequence listing.
- 15 5. A recombinant expression vector containing a transcription unit comprising a DNA sequence according to any one of the preceding claims, a transcriptional promoter, and a polyadenylation sequence.
- 20 6. A host cell transformed with the recombinant expression vector of claim 5.
7. A process for the manufacture of a glucuronyl C5-epimerase or a functional derivative thereof capable of converting D-glucuronic acid (GlcA) to L-iduronic acid (IdoA), comprising cultivation of a cell line transformed 25 with a recombinant expression vector according to claim 5 in a nutrient medium allowing expression and secretion of said epimerase or functional derivative thereof.
8. A glucuronyl C5-epimerase or a functional derivative thereof whenever prepared by the process of claim 7.

1/2

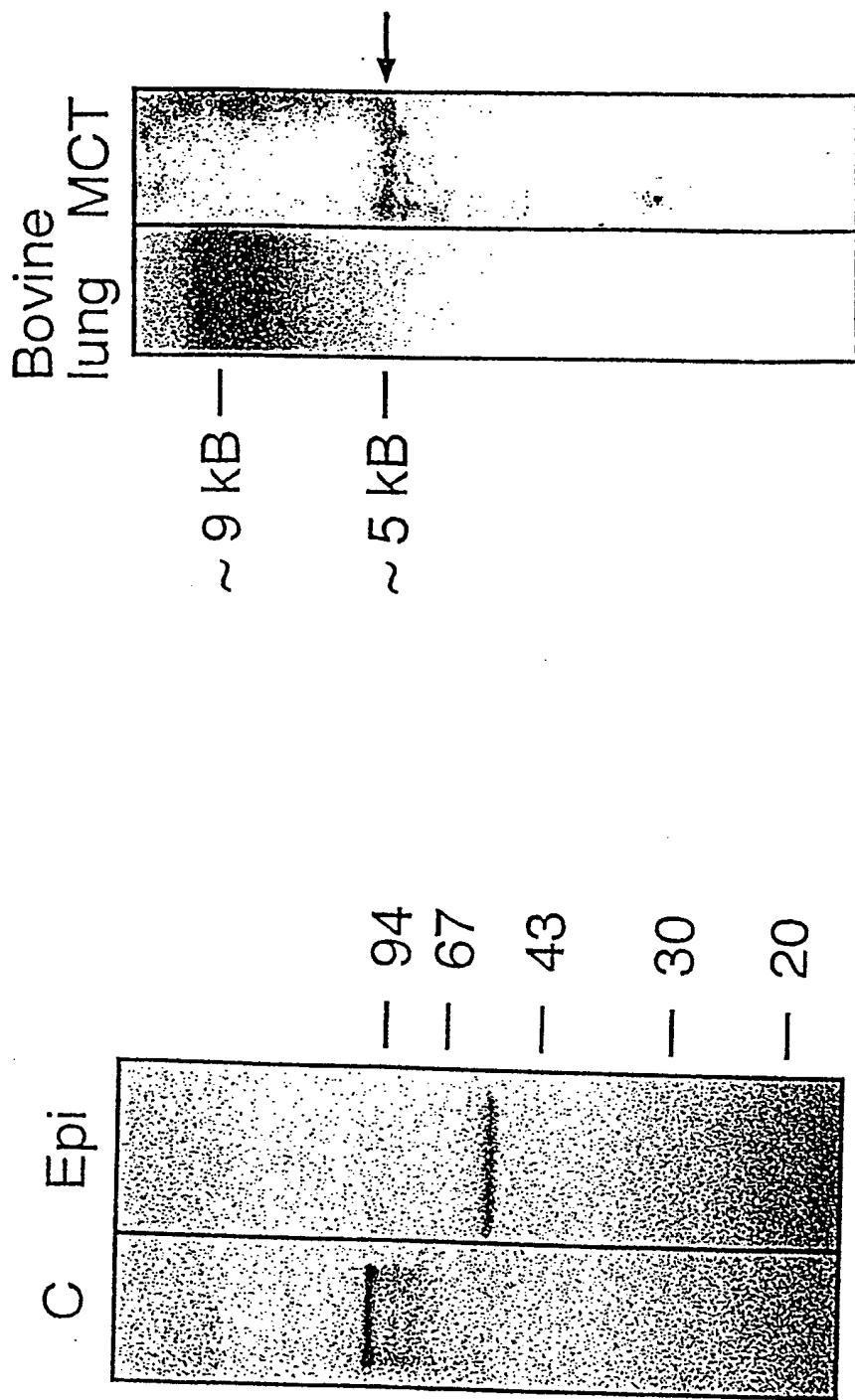


Figure 1

Figure 3

2/2

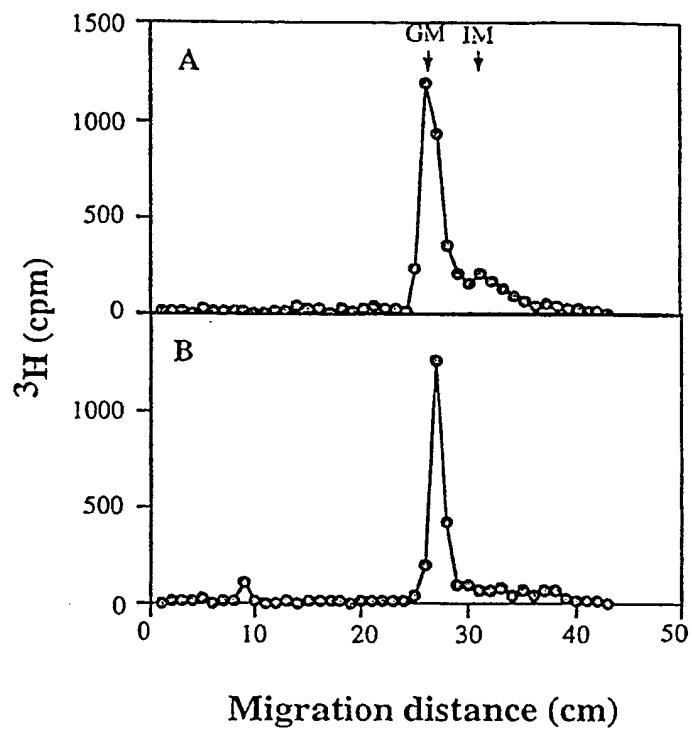


Figure 2

## INTERNATIONAL SEARCH REPORT

International application No. PCT/SE 98/00703
--

## A. CLASSIFICATION OF SUBJECT MATTER

**IPC6: C12N 9/90**

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

**IPC6: C12N**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

**SE,DK,FI,NO classes as above**

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	The Journal of Biological Chemistry, Volume 269, No 43, October 1994, Patrick Campbell et al, "Biosynthesis of Heparin/Heparan Sulfate", page 26953 - page 26958  --	1-8
A	WO 9614425 A1 (INALCO S.P.A.), 17 May 1996 (17.05.96)  -- -----	1-8

 Further documents are listed in the continuation of Box C. See patent family annex.

- \* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search  <b>30 June 1998</b>	Date of mailing of the international search report  <b>02-07-1998</b>
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. + 46 8 666 02 86	Authorized officer  <b>Yvonne Siösteen</b> Telephone No. + 46 8 782 25 00

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

09/06/98

International application No.

PCT/SE 98/00703

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9614425 A1	17/05/96	AU 3926195 A CA 2204366 A EP 0789777 A IT 1271057 B IT MI942240 D	31/05/96 17/05/96 20/08/97 26/05/97 00/00/00